

Same as OKUBO, DNA Res.
1, 37-45 (1994)

世界知的所有権機関

国際事務局

PCT

特許協力条約に基づいて公開された国際出願



<p>(51) 国際特許分類6 C12N 15/11, C12Q 1/68 // G01N 33/566</p>	<p>A1</p>	<p>(11) 国際公開番号 WO 95/14772</p> <p>(43) 国際公開日 1995年6月1日 (01.06.95)</p>
<p>(21) 国際出願番号 PCT/JP94/01916 (22) 国際出願日 1994年11月11日(11.11.94)</p> <p>(30) 優先権データ 特願平5/355504 1993年11月12日(12.11.93) JP</p> <p>(71) 出願人; および (72) 発明者 松原謙一(MATSUBARA, Kenichi)[JP/JP] 〒565 大阪府吹田市山田東3-18-1-804 Osaka, (JP) 大久保公策(OKUBO, Kousaku)[JP/JP] 〒562 大阪府箕面市瀬川2-11-26 Osaka, (JP)</p> <p>(74) 代理人 弁理士 吉田研二, 外(YOSHIDA, Kenji et al.) 〒180 東京都武蔵野市吉祥寺本町1丁目34番12号 Tokyo, (JP)</p>		<p>(81) 指定国 AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, JP, KG, KR, KZ, LK, LR, LT, LV, MD, MG, MN, NO, NZ, PL, RO, RU, SI, SK, TJ, TT, UA, US, UZ, VN, 欧州特許(AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI特許(BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO特許(KE, MW, SD, SZ).</p> <p>添付公開書類 国際調査報告書 補正書</p>
<p>(54) Title : GENE SIGNATURE</p> <p>(54) 発明の名称 ジーン・シグナチャー</p> <p>(57) Abstract</p> <p>A 3'-directed cDNA library which accurately reflects the abundance ratio of mRNA in a cell has been prepared from various human tissues, and sequencing of the cDNAs contained in the library has been conducted to examine the incidence of each cDNA in each tissue. As each cDNA has expression information with each tissue corresponding to the mRNA concentration, these cDNAs are usable as a probe or primer for detecting cell anomaly or discriminating cells. The cloned gene can produce proteins utilizable as a medicine or the like.</p>		

(57) 要約

種々のヒト組織から、mRNAの細胞内の存在割合を忠実に反映する3'指向cDNAライブラリーを作成した。該ライブラリーに含まれるcDNAを配列決定し、組織毎の各cDNAの出現頻度を調べた。各cDNAにはmRNA濃度に対応する組織毎の発現情報が付加されているので、該cDNAは、細胞の異常を検出したり細胞の識別をするためのプローブ・プライマーなどとして用いることができる。またクローニングされた遺伝子は、医薬品などに利用し得る蛋白質を産生可能である。

情報としての用途のみ

PCTに基づいて公開される国際出願をパンフレット第一頁にPCT加盟国を同定するために使用されるコード

AM	アルメニア	EE	エストニア	LK	スリランカ	RU	ロシア連邦
AT	オーストリア	ES	スペイン	LR	リベリア	SD	スーダン
AU	オーストラリア	FI	フィンランド	LT	リトアニア	SE	スウェーデン
BB	バルバドス	FR	フランス	LU	ルクセンブルグ	SG	シンガポール
BE	ベルギー	GA	ガボン	LV	ラトヴィア	SI	スロベニア
BG	ブルガリア	GB	イギリス	MC	モナコ	SK	スロバキア
BJ	ベナン	GE	グルジア	MD	モルドバ	SN	セネガル
BR	ブラジル	GN	ギニア	MG	マダガスカル	SS	スワジランド
BY	ベラルーシ	GR	ギリシャ	ML	マリ	TD	チャド
CC	カナダ	HU	ハンガリー	MN	モンゴル	TG	トーゴ
CF	中央アフリカ共和国	IE	アイルランド	MR	モリタニア	TH	タイ
CG	コンゴ	IS	アイスランド	MW	モザンビーク	TJ	タジキスタン
CH	スイス	IT	イタリア	MX	メキシコ	TM	トルクメニスタン
CI	コート・ジボアール	JP	日本	NE	ニジェール	TT	トリニダード・トバゴ
CM	カメルーン	KE	ケニア	NL	オランダ	UA	ウクライナ
CN	中国	KP	朝鮮民主主義人民共和国	NO	ノルウェー	UG	ウガンダ
CZ	チェコ共和国	KR	大韓民国	NZ	ニュージーランド	US	米国
DE	ドイツ	KZ	カザフスタン	PL	ポーランド	UZ	ウズベキスタン共和国
DK	デンマーク	LI	リヒテンシュタイン	PT	ポルトガル	VN	ベトナム
				RO	ルーマニア		

明細書

発明の名称

ジーン・シグナチャー

技術分野

本願発明は、ヒトジェノミックDNA、ヒトcDNAもしくはヒトmRNAの特定の部位と特異的にハイブリダイズする純化された一本鎖DNAもしくはその相補鎖または該一本鎖DNAとその相補鎖からなる純化された二本鎖DNAに関する。本願発明のDNAは、細胞内の種々のタンパク質に対応するmRNAの発現の状態を総合的または個別的に検出し、疾病、ウイルス感染などによる細胞の異常を検出または診断するため、または細胞の識別・同定をするため、さらには組織特異的に発現している遺伝子等を効率よくクローニングするためなどに用いることができる。本願発明はさらに医薬品などに利用し得る蛋白質を産生できるクローニングされた遺伝子を含むものである。

背景技術

本願発明者らは「細胞の性質は遺伝子の発現パターンの違いによって決定される」というmRNAの最も基本的な属性に着目し、独自のアプローチである「ボディーマッピング (body mapping)」を提唱した。これはヒト体内に約200あるとされている細胞、臓器の「遺伝子の発現情報」を作成、ある遺伝子がいつ、どこで、どの程度発現しているかを明らかにし、遺伝子をそれぞれの発現臓器にマップするという全く新しい試みである。

生体の種々の細胞はその役割に応じて様々なタンパク質を発現しているが、細胞内に存在するタンパク質の濃度は、同一個体内の細胞でも、細胞の種類、発生及び分化の状態、環境等によって異なる。

ところで、遺伝子は、一般に「細胞の生存に必須なタンパク質をコードする遺

伝子」と「細胞の固有の機能を担うタンパク質をコードする遺伝子」とに分類される。このうち、「細胞の生存に必須なタンパク質をコードする遺伝子」は、一般にあらゆる種類の細胞で定常的に発現しており、これらの遺伝子は「ハウスキーピング遺伝子 (housekeeping genes)」と呼ばれている。一方、「細胞の固有の機能を担うタンパク質をコードする遺伝子」は、特定の種類の細胞または特定のグループに属する細胞において特異的に発現していることが多く、また、発生、分化の特定のステージで特異的に発現することもある。また、細胞の置かれた環境によって発現量が増加する「誘発可能な遺伝子 (inducible genes)」であることが多い。これを逆に言えば、「細胞の生存に必須なタンパク質をコードする遺伝子」が発現することによって細胞が生育し、「細胞の固有の機能を担うタンパク質をコードする遺伝子」が発現することによって細胞固有の機能が現出されているといえる。

ところが、疾病、感染等により細胞が異常な状態になると、各細胞内の遺伝子の発現の状態が正常時と比べて変化する。特にウイルスの感染時には、ウイルス固有のタンパク質をコードするRNAが細胞内で大量に合成され、該蛋白が大量に生産される。これを逆に言えば、細胞内の遺伝子の発現レベルの変化、とりわけ細胞内のmRNAの濃度の変化が、細胞を疾病等の異常な状態に導くといえる。

このように、生体内の各細胞の機能と細胞内の遺伝子の発現の状態とは表裏一体の関係にある。従って、分子レベルで各細胞の機能を解明したり疾病の原因を究明したりする際には、細胞内の遺伝子の発現の状態、特に細胞内の各mRNAの濃度を把握することが重要である。

このための1つのアプローチとして、細胞内の全タンパク質を抽出して、その発現の状態を解析することが原理的には考えられる。しかし、特定のタンパク質については単離することが可能な場合もあるが、細胞内では非常に多くの種類のタンパク質が発現しているため、これらのタンパク質を完全に分離し、その発現量を直接測定することは殆ど不可能である。

また、別のアプローチとして、細胞内のタンパク質濃度に対応する細胞内のmRNA濃度を直接測定することが考えられる。しかし、特定のmRNAについては、単離することが可能な場合もあるが、細胞内では非常に多くの種類のmRNAが合成されているうえ、mRNAは抽出の際に酵素による分解を受けやすいなど不安定なものであるため、これらのmRNAを完全に分離し、その合成量を直接測定することはほとんど不可能である。

本願発明は、細胞内の種々のタンパク質に対応するmRNAの発現の状態を総合的または個別的に検出し、疾病、ウイルス感染などによる細胞の異常を検出または診断するため、または細胞の識別・同定をするため、さらには組織特異的に発現している遺伝子等を効率よくクローニングするためのプライマーもしくはプローブなどに用いることができるDNAを提供することを目的とする。さらには医薬品などに利用し得る蛋白質を産生できるクローニングされたDNAを提供することを目的とする。

発明の開示

一般に、遺伝情報の流れはDNA→mRNA→タンパク質の順である（セントラルドグマ、1958年、F. H. C. クリックが提唱）。即ち、DNA上の「タンパク質のアミノ酸配列情報」は、mRNAに一旦転写されてからタンパク質に翻訳される。

更に詳細に述べると、哺乳動物の遺伝子には通常、タンパク質をコードする領域と、遺伝子の発現を制御する領域とが含まれている。遺伝子中のタンパク質をコードする領域は（「エクソン」と呼ばれる）、介在配列（「イントロン」と呼ばれる）によって分断されていることが多い。遺伝子がRNAに転写される際に、前駆体RNA（pre-mRNA）中のイントロンが除去され、エクソンは特定のタンパク質をコードする一連のものとしてタンデムに接続される（これを「スプライシング」という）。一方遺伝子の発現を制御する領域としては、転写領域上流に存在するプロモーター、オペレーターなどの転写を直接制御する領域の他に、転写

される領域内のコード領域の上流（5' 側）および下流（3' 側）に位置する非翻訳領域があげられる。特に、3' 側の非翻訳領域（3' UTR）は、mRNA の輸送および安定性などに寄与しており発現の制御に重要である。pre-mRNA のプロセッシングの過程で、5' 側末端にはキャップ付加が行われ、3' 側の非翻訳領域は特異的な切断を受け、切断部位に通常100～200個のアデニル酸が付加し、ポリ（A）テイルが形成され、またコード領域がスプライシングによって接続されて、mRNA が合成される。合成されたmRNA にリボソームが付加し、タンパク質の合成が行われる。

本願発明者らは、一般に、細胞内の特定のmRNA の存在割合が高いときは、それに対応するタンパク質の発現量も多くなるので、細胞内の各mRNA の存在割合を検出することによって、細胞内の各タンパク質の存在割合を検出することが可能であることを明らかにしている（DNA sequence 2 p. 137-144（1991）、Nature genetics 2 p. 173-179（1992））。

本願発明においても、基本的には常法の如く、逆転写酵素（reverse transcriptase）を用いて、特定の細胞から抽出したmRNA に相補的なcDNAを合成するものであるが、本願発明においては本願発明者らが開発した方法、すなわちmRNA の存在割合が対応するcDNAの存在割合に反映するようにcDNAを合成し、cDNAライブラリーを作成することによって、全mRNAに対応したcDNA群をクローニングし、該ライブラリーを配列決定するという方法を用いる。

なお、本願発明者らのアプローチと一見類似するが、全く異質のアプローチとして、ベンターらのランダムプライミングによって取得したcDNAライブラリーのクローニングが挙げられる。

ベンターらのグループは、市販の脳細胞由来のcDNAライブラリー（米国カリフォルニア州ストラテジーン社、カタログ番号936206、936205又は935）から、ランダムにcDNAをクローニングし、塩基配列を決定してい

る (Science 252 p. 1651-1656 (1991)、Nature 355 p. 632-634 (1992))。

ベンターらの方法は、ランダムプライミングによって得られたcDNAを無作為に配列決定しているが、この方法によれば、

- ①1本のmRNAの様々な領域をランダムにクローニングするので、1本のmRNAから重複する部分がない多くのcDNA断片が生じ、これらのcDNA断片が同一のmRNAに由来するのか、異なるmRNAに由来するのか区別できない。
- ②mRNAが長いほどそのmRNA由来のcDNAが合成される確率が上がる。
- ③ランダムプライマーに含まれる各プライマーの塩基配列によってプライマーとして使用される頻度が異なり、cDNAの合成頻度にばらつきが生じる。

等の理由から、各cDNAの出現頻度は、各mRNAの存在割合に対応しないものになってしまう。従って、ベンターらの方法によっては、仮に無作為のクローニングの際の各cDNAの出現頻度をとったとしても細胞内の各mRNAの存在割合を検出することは不可能であり、細胞内のタンパク質の存在状態を知ることとは不可能である。

しかし、本願発明者らの開発した方法によりDNAライブラリーを作成すれば、上記問題点が全て解決され、mRNAの存在割合を忠実に反映したcDNAライブラリーが作成可能である。本願発明においては、「ポリT」のみをプライマーとしてcDNAを合成している。cDNAの3'末端には「ポリAテイル (poly A tail)」と呼ばれる、Aが連続した構造が存在するので、「ポリT」のみをプライマーとしてcDNAを合成すれば、必ず3'末端からcDNAの合成が開始され、3'末端側のcDNAが合成される。また、mRNAの3'末端の大半がそのmRNA独自のものであり、他の種類のmRNAには存在しないものであるので、(Birnstiel et al., Cell. 41, 349-359 (1985))、3'末端側のcDNAはほとんどが特定のmRNAのみにハイブリダイズするものとなる。cDNAを合成後、制限酵素MboI (4塩基を認識し切断する)でcDNAを切断すれば、3'側から最初のMboI部位までのcDNAが作成される。

なお、本願明細書中においては、このようにしてクローニングした「細胞内のmRNAの存在割合を忠実に反映したcDNAライブラリー」に含まれる各々のcDNAを「ジーンシグナチャー (gene signature) (以下GSと略称する) と称する。ただし、GSには二本鎖DNAばかりでなく、それぞれの一本鎖も含まれる。

本願発明は、配列番号1から7837に記載された塩基配列のいずれかからなる一本鎖DNAもしくはその相補鎖、または該一本鎖DNAもしくはその相補鎖の一部を含み、ヒトジェノミックDNA、ヒトcDNA、もしくはヒトmRNAの特定の部位に特異的にハイブリダイズする純化された一本鎖DNAもしくはその相補鎖またはこれらの一本鎖DNAとその相補鎖からなる純化された二本鎖DNAに関する。また、本願発明は、該一本鎖DNAからなるプローブ、プライマーに関する。更に本願発明は、「配列番号1から7837に記載された塩基配列 (ただしTをUと読み変える) のいずれかもしくは該塩基配列の一部を3'側に含むヒトmRNA」に相補的な一本鎖DNAもしくはその相補鎖、または該一本鎖DNAもしくはその相補鎖の一部を含み、ヒトジェノミックDNA、ヒトcDNA、若しくはヒトmRNAの特定の部位に特異的にハイブリダイズする純化された一本鎖DNAもしくはその相補鎖、またはこれらの一本鎖DNAとその相補鎖からなる純化された二本鎖DNAに関する。また、本願発明は一本鎖DNAからなるプローブ、プライマーに関する。

以下、本願発明をより詳細に説明する。

本願発明のDNAは、配列番号1から7837に記載された塩基配列のいずれかからなる一本鎖DNAもしくはその相補鎖ばかりではなく、ヒトジェノミックDNA、ヒトcDNA又はヒトmRNAとハイブリダイズすれば配列番号1から7837に記載された塩基配列のいずれかからなる一本鎖DNAもしくはその相補鎖の一部を「含む」一本鎖DNAであってもよい。

さらに、本願発明のDNAは、「配列番号1から7837に記載された塩基配列 (ただしTをUと読み替える) のいずれかもしくは該塩基配列の一部を3'側

に含むヒトmRNA」に相補的な一本鎖DNAもしくはその相補鎖ばかりではなく、該一本鎖DNAもしくはその相補鎖の一部を含み、ヒトジェノミックDNA、ヒトcDNAまたはヒト、RNAとハイブリダイズする一本鎖DNAもしくはその相補鎖の一部を含む一本鎖DNAであってもよい。

さらに、それらのDNAは一本鎖のものばかりではなく、その相補鎖または、該一本鎖DNAとその相補鎖からなる二本鎖DNAも含まれる。

なお、ここでいう「含む」とは①「各配列番号に記載されている塩基配列のいずれかからなる一本鎖DNAもしくはその相補鎖、またはそれらの一部」または②「各配列番号に記載された塩基配列（ただしTをUと読み替える）のいずれかもしくは該塩基配列の一部を3'側に含むヒトmRNAに相補的な一本鎖DNAもしくはその相補鎖、またはそれらの一部」が、本願発明のDNAに「連続して含まれている」場所が一ヶ所であることのみを表わすものでないことは明白である。すなわち、「含む」とは、本願発明のDNAに存在する前記①の塩基配列または前記②の塩基配列の間に、他の塩基の挿入が存在する場合等も該当することは明らかである。

なお、ヒトジェノミックDNA、ヒトcDNAまたはヒトmRNAの特定の部位とのハイブリダイズの条件は、通常行われている条件を適用することができる（例えば、Molecular Cloning: A Laboratory Manual、Sambrookら、Cold Spring Harbor Laboratory Press、1989）。

次に、各細胞内のmRNAの存在割合を忠実に反映するcDNAライブラリーの作成方法および全mRNAに対応したcDNA群のクローニング方法、ならびに各cDNAの塩基配列の決定方法についての一例を下記に記載する。

まず、特定の組織に由来する細胞、例えば臓器由来の細胞、例えば、ヒト肝臓由来細胞（HepG2）を培養し、常法に従って、全mRNAを単離する。次いで、得られたmRNAをcDNAライブラリー作成のためのベクターに挿入する。

例えばpUC19（クローニング部位側にM13の配列を有する）には、下記

の方法で挿入する。

すなわち、pUC19をHincIIおよびPstIで開裂し、PstI切断末端に20bpから30のbp程度のポリTを付加し、そこにmRNAの3'側のポリAをハイブリダイズさせる(図1a)。そして、常法により逆転写酵素を用いてDNA鎖を伸長させた後、例えばDNAポリメラーゼにより二本鎖DNAを合成する(図1b)。得られた二本鎖DNAを4塩基を認識する制限酵素MboIで切断する(図1c)。

MboIは4塩基認識の制限酵素であるから、挿入cDNAのポリAテールから数百塩基程度離れた位置でDNAを切断すると考えられる。そしてMboIは、本願発明者らがGenBankのデータベースから無作為に抽出した約300のヒトcDNAを例外無く切断することを確認したことから、MboIはクローニングされるcDNAのいずれかの部位を切断すると考えられる。なお、pUC19はdam⁺の大腸菌、例えばE. coli JM109から調製するので、MboI認識部位のアデニンがメチル化されており(G^mATCとなっており)、MboIによっては切断されない。

次に、pUC19に導入されたMboI切断部位を有する二本鎖DNAをベクター内に組み込むために、ベクター側にMboIと同じ粘着末端を生ずるBamHIを作用させる。この場合に、BamHI(GGATCC)は、MboI(GATC)の認識塩基配列を含むため(GGATCC)、伸長された二本鎖DNAが、更に当該BamHIによって切断されることはない。

その後、常法により上記二本鎖DNAをライゲーションし、組み換え体プラスミドを大腸菌例えばE. coli DH5へ導入してcDNAライブラリーを作成する。

この方法によれば、mRNAの3'側ポリA上流の配列のみを含むクローンを取得することができる。

また、得られるcDNA断片の長さが平均270bpと短いため、cDNA断片が長すぎることから生ずるcDNA合成効率の差や形質転換時の効率の差から